Low Back and Common Widespread Pain Share Common Genetic Determinants

Ida Malkin1, Frances M. K. Williams2, Genevieve LaChance2, Timothy Spector2, Alex J. MacGregor2,3 and Gregory Livshits1,2

1Department of Anatomy and Anthropology, Sackler Faculty of Medicine, Tel Aviv University, Israel
2Department of Twin Research and Genetic Epidemiology, King's College London, UK
3School of Medicine, University of East Anglia, Norwich, UK

Summary

Low back (LBP) and chronic widespread musculoskeletal pain (CWP) both have a significant genetic component and are associated with increased body mass index (BMI). We examined whether LBP and CWP share common genetic factors, and to what extent this correlation is modified by the genetic factors influencing BMI. Genetic analysis of binary traits such as pain is not simple, particularly if their risk is associated with age or other quantitative traits. Implementing Falconer’s polygenic threshold concept for dichotomous traits inheritance, we developed new software to examine the extent of the genetic influence on LBP and CWP under age and BMI dependence. The analysis was conducted on 3266 and 2256 UK female twins, assessed for LBP and CWP, respectively. Analysis of the liability scores with threshold to LBP and CWP established substantial contribution of genetic factors to their variation ($h^2 > 0.60$, $p < 0.004-0.0003$) and covariation ($p = 3.1E-08$). Some 39% of the CWP and 70% of the LBP heritability estimates were attributable to genetic effects shared by both phenotypes, and 40% and 67% of the residual variation is caused by environmental factors simultaneously affecting both pain syndromes. However, contribution of BMI to variation/covariation of both pain phenotypes—although statistically highly significant ($p < 10^{-7}$)—was not determinative.

Keywords: LBP, CWP, threshold, liability score, heritability, genetic correlation

Introduction

Episodes of acute low back pain (LBP) and chronic widespread musculoskeletal pain (CWP) are both common clinical musculoskeletal conditions (Papageorgiou et al., 2002; Rubin, 2007; Gerdle et al., 2008). Despite many years of research, the etiology of both pain conditions remains poorly understood: a range of behavioral, dietary, mechanical, and neuroendocrine abnormalities have been explored (Macfarlane et al., 2009; Vandenkerkhof et al., 2011). Family and twin studies of CWP have shown that genetic factors play a major role in the development of both conditions, with heritability estimated to be in the vicinity of 50% (MacGregor et al., 2004; Kato et al., 2006; Battié et al., 2007). One of the most consistent features of epidemiological studies of CWP and LBP has been the documented association of both pain phenotypes with raised body mass index (BMI). This has been demonstrated both in cross-sectional studies (Neumann et al., 2008; McCarthy et al., 2009; Heuch et al., 2010) and, more recently, in longitudinal work (Mork et al., 2010; Heuch et al., 2013). The longitudinal design is advantageous in that it provides evidence that BMI is raised prior to the onset of pain, rather than weight gain resulting from, for example, pain-induced immobility. These studies suggest, therefore, that raised BMI is a risk factor for developing both CWP and LBP.

Since both conditions manifest with chronic pain, we postulated that they share an underlying genetic predisposition, which is at least partly related to BMI. The bivariate analysis of twins allows examination of the genetic and environmental determinants of the two conditions and estimation of the degree to which the genetic predisposition is shared. However, the quantitative genetic analysis of binary variables (yes vs no) such as pain is not a simple task, in particular if their
manifestation demonstrates age and/or other quantitative trait dependence. In the present study, we developed new software to examine the extent of the genetic effects (heritability) on the presence of age-dependent binary traits. These methods were used to examine the genetic covariation between liability to pain syndromes and the risk factor BMI.

Materials and Methods
Study Sample and Phenotypes
For CWP, a total of 3266 Caucasian females (age range 18–84 years, with mean 54.4 years) were examined in this project. These comprised 756 and 696 pairs of monozygotic (MZ) and dizygotic (DZ) twin pairs, respectively, and 362 singletons (twins without sibling measurement). In the present study, twins with CWP information had completed standardized questionnaires following a modified version of the London Fibromyalgia Symptom Screening questionnaire (LFESSQ) (White et al., 1999). Twins diagnosed using this method have been included in a recent genome-wide association meta-analysis of CWP (Peters et al., 2013). Individual twins were considered as cases if they reported having pain on both sides of the body, above and below the diaphragm, with duration of 7 days or more within the preceding 3 months. In addition, 2256 twins (371 MZ and 698 DZ pairs) had completed at home a standardized questionnaire without conferring, relating to lifetime history of low back symptoms, using a modified version of the MRC Nurses Study (Smedley et al., 1998) and included both questions and a mannequin. LBP was defined as being located between the 12th ribs and the gluteal folds. Pain associated with fever, menstruation, or pregnancy was excluded. Our analyses focused on pain associated with disability in activities of daily living, defined as any one of the following being impossible (as opposed to just “difficult”): walking around the house, standing for 15 min, getting up from a low chair, getting out of the bath, getting in and out of the car, going up and down stairs, putting on socks or stockings, and cutting toenails. Further detailed description is given elsewhere (MacGregor et al., 2004; Livshits et al., 2011).

Both the CWP and the LBP phenotypes were defined as a binary traits (e.g., 1 = affected vs 0 = nonaffected). The participants had undergone height and weight measurements used to calculate BMI. St Thomas’ Hospital ethics committee approval had been obtained and the participants gave written informed consent.

Statistical–Genetic Analysis
First, using STATISTICA (www.statsoft.com) we computed preliminary descriptive statistics, and compared affected and nonaffected individuals. We also computed tetrachoric correlation coefficients by zygosity to evaluate whether the pain phenotype aggregated within the families.

Quantitative genetic analysis was based on Falconer’s concept of heritability of threshold characters (Falconer, 1965; Falconer & Mackay, 1996). Briefly, this approach assumes the existence of an unobserved, normally distributed liability to the disease caused by the independent effects of additive genetic and environmental factors. The liability determines the threshold above which individuals are considered affected. In a random sample of unrelated individuals, the threshold model is mathematically equivalent to a probit-like risk model (Curnow, 1972). That is, if the observed binary trait phenotype (B) is related to the continuous liability (X) by way of a threshold (τ), then the probability that an individual is affected: P(Bn = 1) = P(Xn > τ). Assuming each individual in the sample has some covariate measurements Yn, the probability density of individual liability can be presented as a normal distribution around the predicted value F(Yn). The multiple regression coefficients for covariates in F(Y) are determined in maximum likelihood estimates (MLE). Following Falconer (1965), we assumed in our analysis a multifactorial, polygenic nature of the liability scores. This allows implementation of a variance component analysis (VCA), assuming that variation in liability scores is caused by orthogonal (independent) variance components, namely, V X = V AD + V CE + V RS, where V AD, V CE, V RS are variance components reflecting additive genetic, common family, and random environment effects, respectively (Falconer & Mackay, 1996).

Consider a pedigree including N individuals. Each individual has a unique predicted value of liability F(Yn). The probability density of joint pedigree liability residuals can be expressed as N-variable normal distribution with nonzero correlation between the relatives. The probability of the dichotomous trait values set {Bn} (affection status) observed in the pedigree members is computed as an N-dimensional threshold integral of the joint probability density of each individual variable Xn, from -∞ to τ (if Bn = 0), or from τ to +∞ (if Bn = 1). The computation of this integral is an immense task, and therefore here we propose a new formulation of joint VC likelihood, which makes this computation easier (see Appendix). We propose modified quasi VCA (QVCA) joint likelihood expression to fit the liability transmission and a QVCA maximization procedure estimating the variance of components (V AD, V CE, V RS) and the affection threshold τ on a liability scale. A major strength of this approach is that our software allows the explicit inclusion of covariates in the model and estimates proportion of liability variance attributable to covariates. The procedure is written on C++ and included in the new version of MAN package for pedigree analysis (Malkin & Ginsburg, 2014).
Using the likelihood ratio test (LRT) as a model-fitting technique, a best fitting and most parsimonious model was obtained for the pattern of inheritance of CWP, LBP, and BMI.

Since significant associations were observed between the study phenotypes, we examined to what extent these associations could be attributed to shared genetic or environmental factors. To test this hypothesis, we used bivariate QVCA, which assumes that observed phenotypic correlations can be caused by pleiotropic genetic factors (as measured by additive genetic correlation), and/or by shared environmental factors (measured by environmental correlation) (Falconer & Mackay, 1996).

**Results**

Table 1 provides the basic descriptive statistics of the study sample. Since LBP and CWP were assessed in TwinsUK at different time-points and were not fully overlapping (1297 individuals had completed both pain questionnaires) the corresponding anthropometric and demographic information is given separately. Cases of both pain phenotypes showed the same trend compared to controls: they were on average older and heavier in body weight. The \( \chi^2 \) test for the independence of these two phenotypes (Table 2) rejected the null hypothesis of no association (\( P < 0.0001 \)). Next using the binary logistic regression, we examined whether the probability of an individual having LBP or CWP was associated with increased BMI, independently of age. In both cases, the results were statistically significant with odds ratios (OR) per SD of BMI change: 1.325 (1.200–1.462, \( p = 2.3\times10^{-8} \)) and 1.357 (1.225–1.503, \( p = 4.6\times10^{-9} \)).

We then examined the contribution of common and trait specific genetic and environmental factors to inter-individual variability and co-variability of liability score and LBP, CWP, as well as BMI. Tetrachoric correlations for MZ vs DZ twins by phenotype were: 0.733 vs 0.422 for CWP, 0.634 vs 0.363 for LBP, and 0.526 vs 0.179 for cross-phenotype associations, and were statistically significant: \( P = 0.015 \) for DZ twins cross-phenotype associations, and \( P < 0.001 \) for all others. These estimates provide evidence of the familial, likely genetic, factors contributing to the risk of chronic pain. This was confirmed in a univariate modeling of CWP and LBP variations (Table 3). In particular, this shows that the contribution of common twin environment was negligible (\( P > 0.20 \)), while contribution of the putative genetic factors was quite substantial (\( h^2 > 0.60 \)) and by LRT statistically highly significant with \( p = 4.2\times10^{-3} \) and 2.6\times10^{-4} for LBP and CWP, respectively. The threshold for the corresponding liability scores for LBP and CWP was estimated at 0.72 SD and 0.84 SD, respectively.

The bivariate model also showed a highly significant association between LBP and CWP caused both by shared genetic and shared environmental factors, resulting in significant genetic (\( R_{AD} \)) and environmental (\( R_{RE} \)) correlations (Table 4). Parameters \( v_{AD}[1], v_{AD}[2] \) determine for each trait the proportion of additive genetic variance attributed to genetic factor, which influences both traits simultaneously; \( v_{RS}[1], v_{RS}[2] \) estimate the same proportions for individual environmental variance (see Appendix). These parameters indicate that 39% for CWP to 70% for LBP of the trait heritability is attributable to shared genetic effects, and roughly 40% and 67% of the residual variation is caused by the environmental factors simultaneously affecting both pain syndromes.

| Table 1 | Basic descriptive statistics of the study sample by affection status. |
|---|---|---|---|---|---|---|---|---|---|
| Trait | LBP = 0 (1710) | LBP = 1 (546) | CWP = 0 (2585) | CWP = 1 (681) |
| Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| Age [years] | 49.95 | 11.78 | 51.48 | 10.28 | 2.92 | 53.27 | 14.21 | 58.77 | 10.94 | 10.91 |
| Height [m] | 1.62 | 0.06 | 1.63 | 0.06 | 3.39 | 1.62 | 0.06 | 1.61 | 0.06 | 4.60 |
| Weight [kg] | 64.89 | 11.32 | 68.97 | 12.93 | 6.60 | 66.41 | 12.03 | 70.95 | 14.40 | 7.57 |
| BMI [kg/m²] | 24.69 | 4.29 | 25.92 | 4.79 | 5.35 | 24.48 | 4.30 | 26.37 | 5.20 | 8.75 |

* t-tests compare the corresponding variable between the affected (1) and unaffected (0) individuals, for each pain phenotype separately, all tests have \( P < 0.05 \).

| Table 2 | Affected vs unaffected CWP and LBP individuals in study sample (2×2 contingency table). The sample shows statistically highly significant dependence (coincidence) between the two pain phenotypes \( \chi^2 = 107.4, P < 0.0001 \). |
|---|---|---|---|---|
| CWP \( \neq 1 \) | CWP = 0 |
| LBP/CWP | Expected | Observed | Expected | Observed | Total |
| LBP = 1 | 72.6 | 140 | 225.4 | 158 | 298 |
| LBP = 0 | 243.4 | 176 | 755.6 | 823 | 999 |
| Total | 316 | 316 | 981 | 981 | 1297 |

* 1 and 0 represent affected and unaffected individuals correspondingly.

© 2014 John Wiley & Sons Ltd/University College London
Table 3  Best fitting and most parsimonious models for BMI and for liability scores to LBP and CWP.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>BMI</th>
<th>LBP</th>
<th>CWP</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_{AD}$</td>
<td>0.710 ± 0.043</td>
<td>0.624 ± 0.218</td>
<td>0.636 ± 0.174</td>
</tr>
<tr>
<td>$V_{SB}$</td>
<td>[0]</td>
<td>[0]</td>
<td>[0]</td>
</tr>
<tr>
<td>$V_{RS}$</td>
<td>{0.261}</td>
<td>{0.362}</td>
<td>{0.257}</td>
</tr>
<tr>
<td>$a_0$</td>
<td>0.099 ± 0.020</td>
<td>0.098 ± 0.043</td>
<td>0.215 ± 0.035</td>
</tr>
<tr>
<td>$b_1[\text{Age}_s]$</td>
<td>0.166 ± 0.016</td>
<td>0.079 ± 0.038</td>
<td>0.129 ± 0.027</td>
</tr>
<tr>
<td>$b_2[(\text{Age}_s)^2]$</td>
<td>-0.036 ± 0.013</td>
<td>-0.079 ± 0.038</td>
<td>-0.129 ± 0.027</td>
</tr>
<tr>
<td>$\tau$</td>
<td>0.716 ± 0.035</td>
<td>0.839 ± 0.030</td>
<td>0.107</td>
</tr>
<tr>
<td>$V_{age}$</td>
<td>0.029</td>
<td>0.014</td>
<td>0.107</td>
</tr>
<tr>
<td>$P{\text{LRT}(V_{AD}=0)}$</td>
<td>&lt;1.0E-17</td>
<td>4.2E-03</td>
<td>2.6E-04</td>
</tr>
</tbody>
</table>

$a, b$ are regression parameters reflecting effect of age and age-squared; $\tau$ – affection threshold of liability; $V_{age}$ proportion of variance attributed to age; $P\{\text{LRT}(V_{AD}=0)\}$ - $P$-value of likelihood ratio test for significance.

Next we tested the extent to which the association between each pain variable with BMI might be explained by common genetic and environmental factors. Bivariate analyses of the pain traits with BMI revealed that for both variables the genetic correlations ($R_{AD}$) were modest, 0.205 ± 0.015 for LBP and 0.185 ± 0.013 for CWP, but highly significant ($p<10^{-17}$, Table 4). The environmental correlation between CWP and BMI was significant ($R_{RS}=0.243 ± 0.065, p=1.8E-04$). To estimate the extent of genetic and/or environmental correlation between LBP and CWP that were attributable to the association of each trait with BMI, we added BMI as a covariate to the bivariate analysis between LBP and CWP. As expected, BMI appeared to be a significant covariate ($p=8.7E-05$ and $7.21E-11$, respectively, by LRT) to liability score variation of each of the pain phenotypes. However, the genetic correlation between them diminished to $R_G = 0.502 ± 0.096$ vs $0.525 ± 0.095$, observed prior to adjustment for BMI.

Discussion

The major aim of this study was to assess the genetic and environmental risk factors contributing to LBP and CWP in females and to define more precisely the role of BMI, which is a known risk factor for both conditions. Both pain phenotypes were associated with age, and the development of software which can take such covariates into account is a major advantage. We have shown for the first time that two musculoskeletal pain traits, complaints which are highly prevalent in general practice (Hensler et al., 2009) and disability adjusted for life years (Murray et al., 2012), are linked by shared genetic factors. Specifically, parameterizing CWP and LBP in a group of twins onto an unobserved continuous variable, liability scale, and accounting for the family resemblance, we showed that some 60% of the liability score variation for each of these traits is attributable to genetic factors. This is also in good agreement with the tetrachoric correlation
representing the correlation between two hypothetical normally distributed traits, with predetermined threshold. This could be considered as the simplest estimate of liability score correlation, if we do not require the equality of the affection threshold, \( \tau \) in MZ and DZ twins, and covariates effect is not taken into account. Under these limitations, the observed tetrachoric correlations yield heritability estimates 0.58 and 0.54 for CWP and LBP, respectively.

Our estimates are somewhat higher than those obtained in other samples, although in some studies the estimates are quite similar to the present ones. For example, a recent review paper (Nielsen et al., 2012) summarizing most of the data published on LBP heritability noted the wide range of heritability estimates, from 0% to 68%. Two major factors may be responsible: a heterogeneous definition of LBP phenotype and huge heterogeneity in age of the studied samples, with lowest \( h^2 \) estimates in pediatric and geriatric age groups. CWP examinations in females also give substantial estimates of heritability, in the vicinity of 50% (Nielsen et al., 2012).

Our method, testing the heritability on liability scale has a number of advantages, since the dependent variable has all the features of a quantitative continuous trait. It allows adjustment for covariates, such as age or body composition, directly in the model and can compute genetic and environmental correlations, similar to classical bivariate variance decomposition analysis. There are several pieces of software for estimating the heritability of the dichotomous variables. For example, the Mx scripts library (Posthuma & Boomsma, 2005) includes procedures to test heritability of ordinal traits, implementing heritability estimates based on classical ACE modelling in samples of MZ and DZ twin pairs. However, covariates cannot be included in the computation. Another Mx script designed for raw ordinal data in MZ and DZ twin samples allows adjustment for age and has been used to fit an ACE model for cognitive dysfunction (Reynolds et al., 2006).

In this model, liability for all individuals is distributed with zero mean, but liability affection threshold is age-dependent based on a regression model including age and age\(^2\), and conducted by sex separately. In our procedure, we assumed the liability threshold is the same for all individuals, but individual liability predicted by covariates is included as distribution mean (as is usually done for quantitative traits). Predicted liability can include up to five covariates; for each covariate the multiple regression coefficients can be sex specific or equal for both sexes. Significance of each covariate and each variance component can be tested by LRT to produce the most parsimonious model.

This flexibility of modelling allowed us to conduct a bivariate analysis with adjustment for age and BMI, and to show that the variations of the CWP and LBP are not independent, as could be judged from the significant estimates of the genetic and environmental correlations. The common factors explain about 40% and 70% of the genetic and environmental effects, respectively, thus suggesting very close etiology of the two traits. On the other hand, despite highly significant association of both pain phenotypes with BMI (Table 1), and statistically extremely significant (\( P \approx 10^{-17} \)) genetic correlations (Table 4), the magnitude of these correlations was low, and explained only a minor part of LBP/CWP association. This is useful information for the chronic pain field: it suggests that genetic and metabolic pathways play a major role in governing LBP and CWP separate from those influencing BMI.

Several attempts to estimate the extent of the common genetic factors on different pain phenotypes have been undertaken previously. For example, we have previously implemented classical twin analysis estimated heritability of the pain scores at seven skeletal sites (such as neck, back, elbow, etc.) in TwinsUK (Williams et al., 2010). Multivariate analysis of these data concluded that 95% of the variation of these pain phenotypes could be attributable to a single common factor, with estimated heritability of 0.46. The estimates of heritability of lumbar, thoracic, and neck pain were between 0.30 and 0.40, with genetic correlations varying between the ~0.35 and 1.00, in a very large sample (>15,000 individuals) of Danish twins (Hartvigsen et al., 2009). The conclusion was similar to Williams et al. (2010) that there is likely a common genetic basis for the pain at various sites of the spine.

While phenotype definition is clearly important in genetic epidemiological studies, there is evidence that grouping of similar syndromes allows for greater power to identify association with the specific genetic factors. For example, shared genetic variants have been shown to predispose to several types of major psychosis (Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013). Substantial and highly significant overlapping of the genetic factors affecting variation of several correlated phenotypes, as was observed in this and other studies (Williams et al., 2010), e.g., anxiety disorders (Tambs et al., 2009) and psychotic phenotypes mentioned above, may suggest the existence of an underlying common endophenotype. This has important consequences in both fundamental science, for example, in genome-wide association study design, and applied science, such as pharmacogenetic development of novel analgesics. Covariates such as BMI with highly significant association as observed in the present study, but also by others (e.g., Hershkovich et al., 2013), may have little practical gene-finding significance due to their modest contribution to risk of the disease manifestation.

There are several limitations to the present study. The main one includes two time-points of questionnaires leading to unequal sample sizes, and overlap of only 1297 twins. It is also important to mention that some overlap between both pain syndromes is also possible. It is, for example, possible that LBP...
may spread to become CWP. Some bias is possible because the definitions of both syndromes are based on subjective perception. Nevertheless, it should be stressed that we have used standardized definitions for both conditions, widely used in this field of study, and in practice there is little chance of confusing the two in a cross-sectional setting. The other limitation of the study is that data on the lifestyle of these subjects, which may be of importance as risk factors in chronic pain, were not available. The study also focused on females, and the conclusions may not be fully applicable to men. However, there are advantages in a study of women: both LBP and CWP are more prevalent in women, and genotype * sex interaction in complex phenotypes is well recognized (Vink et al., 2012). Our data show little evidence of MZ vs DZ difference in diagnosis and mis-assignment, which would serve to reduce the genetic effects observed. There was no evidence of zygosity-related difference in any other aspect of the study. The diagnostic questionnaires used in this study were validated and published previously (White et al., 1999; MacGregor et al., 2004). Finally, the implemented statistical-genetic method of analysis allows direct examination of the inheritance and co-inheritance pattern of binary traits, such as pain. Benchek and Morris (2013) recently tested the robustness of heritability estimates of liability to disease, based on twin samples and the ACE model. They concluded that non-normality of the distribution of the environmental component was one of the main sources of biased heritability estimates. From this point of view, the direct inclusion of the covariates, the distributions of which often deviate from normality assumptions (e.g., age, categorical covariates defining life style), is preferable as it can diminish the bias of the genetic effects estimates. Our model also enables simultaneous testing of SNPs and other polymorphisms on dichotomous phenotype liability and can estimate the proportion of variance attributed to individual SNP influence. Thus development of this method paves the way for investigation of numerous other complex phenotypes, with categorical definition of the status.

Acknowledgements

This study was supported by the Israel Science Foundation (grant numbers #994/10, #1018/13). The authors would like to thank all the twins who participated in the study, as well as staff in the Department of Twin Research and would like to acknowledge financial support from the Wellcome Trust, the Pain Research Foundation, and the Department of Health via the National Institute for Health Research (NIHR) comprehensive Biomedical Research Centre award to Guy’s & St Thomas’ NHS Foundation Trust in partnership with King’s College London.

References


If the quantitative trait \( \{X_n\} \) is the hypothetic liability with an affection threshold \( \tau \) relating to a dichotomous affection status \( \{B_n\} \), measured on \( N \) pedigree members, the estimation of model parameters demand computation of \( N \)-dimensional integral of the probability density \( LH \) of the liability, on each variable \( X_n \) from \(-\infty \) to \( \tau \), if \( B_n = 0 \), and from \( \tau \) to \( +\infty \), if \( B_n = 1 \). This is a complex task and is the major reason why VC estimations of disease liabilities have been performed only on samples with simple familial structures including MZ and DZ twin pairs and, importantly, without the inclusion of covariates (Reynolds et al., 2006). Without covariates the two-dimensional integral could be computed numerically for each parameter set only for three co-affection variants of co-twins \([0, 0], [0, 1], [1, 1]\). If covariates are included, the LH for each twin pair in the sample should be integrated separately and the computation time grows unlimitedly with sample size. Unfeasibility to account directly for covariates is a particularly important limitation in the analysis of age-dependent dichotomous phenotypes, including chronic pain phenotypes CWP and LBP. Additionally, deviation from normality assumption of the distribution of the environmental factors produces bias in VC estimates (Benchek & Morris, 2013). Many important covariates including age and categorical lifestyle variables are not normally distributed. Therefore, if these covariates are not included in the model explicitly, the normality of the hypothetical environment component would likely be invalid.

To address this, we have reformulated the VCA LH for quantitative traits, to simplify the integration of probability density \( LH \) on variables \( X_n \) thus enabling the inclusion of several covariates in the analysis of a liability. As a simple illustration, consider a pedigree with two measured individuals \( \{X_{1k}, X_{2k}\} \), both with covariates \( \{Y_{1k}, Y_{2k}\} \) (where \( k \) is the number of the pedigree). The variance of \( X \) in the sample is \( \sigma_X^2 \). To fit the data, we will use a VC model that includes only one variance component \( \sigma_C^2 \) shared by both relatives. \( \sigma_C^2 \) may reflect any type of effect, genetic, environmental, or a combination. Trait value \( X \) can be considered as the sum of independent (orthogonal) traits \( T_j \) (factors): \( X_{1k} = T_{11k} + T_{21k} + T_{31k} = a_k + F(Y_{1k}) + s_{ik}; T_{11k} = a_k \) is a factor specific for the pedigree. Both individuals belonging to the same pedigree \( k \) have the same value \( T_{11k} = T_{12k} = a_k \). The variance of the factor \( T_j \) in the sample is \( \sigma_{C_j}^2 \). \( T_{21k} = F(Y_{2k}) \) is the factor presenting the individual trait value, predicted by covariates. The variance of factor \( T_{2} \) in the sample can be estimated for any function \( F(Y_{ik}) \) and is \( \sigma_{Covar}^2 \). \( T_{31k} = s_{ik} \) is also individual specific \( s_{ik} = X_{1k} - T_{11k} - T_{21k} \). Because all factors are supposed to be orthogonal the variance of \( s_{ik} \) is \( \sigma_{R}^2 = \sigma_{Tot}^2 - \sigma_{Covar}^2 - \sigma_C^2 \).

Let us suppose that \( a_k \) and \( s_{ik} \) are normally distributed in the sample with mean 0 and variance \( \sigma_C^2 \) and \( \sigma_R^2 \) correspondingly. So the probability density \( LH \) for a pedigree can

Appendix

In variance component (VC) analysis, the variance of quantitative trait \( X \) is treated as the sum of normally distributed orthogonal components, associated with different sources of variability. In population biology, these components typically include genetic variability, common environment components (such as common sibling, twin or household components), and an individual specific variability component. For a pedigree, including \( N \) individuals, the standard approach proposes that the likelihood (LH) is formulated as a probability density of \( N \)-variable normal distribution for \( \{X_n\} \) with correlation matrix including nondiagonal items \( r_{n1,n2} \) defined by VCs, shared by individuals \( n1 \) and \( n2 \), and in general with nonzero different individual mean values, predicted by known individual covariates \( \{Y_{ik}\} \).

Despite the fact that VCA was originally proposed for quantitative continuous traits, the approach can be extended to the analysis of categorical traits, as proposed by Falconer (1965). This modified approach assumes the existence of an unobserved, continuously distributed, quantitative variable, the liability. The trait liability is assumed to be normally distributed with a threshold \( \tau \), dividing the population into affected \( (X > \tau) \) and nonaffected \( (X \leq \tau) \) individuals.


be formulated as follows:

\[ LH = 1 / \left( \sigma_C \sigma_R^2 (2\pi)^{1/2} \right) \int_{-\infty}^{+\infty} \exp \left( -\frac{a^2}{2\sigma_R^2} \right) \times \prod_i \exp \left\{ -\frac{(X_i - F(y_i) - a)^2}{2\sigma_R^2} \right\} da. \]  

(1)

Now let us substitute the continuous normal distribution of \( a \), with a stepwise one, having constant probabilities \( p_j \), defined on the intervals having mean values \( \delta_j = j \cdot \beta \), \( -J < j < J \). The sum of all \( p_j \) is equal to 1 and \( p_j = p_{-j} \). The parameter \( \beta \) is selected to produce the same variance \( \sigma_C^2 \):

\[ \sigma_C^2 = \frac{\sigma_R^2}{\sqrt{\sum_{j=-J}^{J} (j/\beta)^2}} \]

The integer number \( J \) and set of \( p_j \) should approximate the distribution of \( a \). The LH for this model is now:

\[ LH = 1 / \left( 2\pi \sigma_R^2 \right) \times \sum_{-J}^{J} \left\{ p_j \prod_i \exp \left\{ -\frac{(X_i - F(y_i) - j\beta)^2}{2\sigma_R^2} \right\} \right\}. \]  

(2)

This form of LH can be extended on pedigrees including \( N \) measured individuals and having \( k \) factors with variances \( \sigma_{C,k}^2 \). Each factor corresponds to an independent variability source, shared for definite \( k \)-th group of individuals in the pedigree. For example, the pedigree, consisting of two parents and one offspring, may be described with four shared factors. Two additive genetic factors each with variance equal to \( 1/2\sigma_{Add}^2 \): the first is shared by father and offspring, the second is shared by mother and offspring. Two environment factors, attributed to spouses and household environment, are shared for parents and for the whole nuclear family correspondingly. Here we have 4 overlapped groups of individuals: 1-father and offspring, 2-mother and offspring, 3-parents, 4-offspring and both parents. The shifts \( j\beta_k \) are included in the exponent only for those individuals who share the appropriate factor. In general for each pedigree member \( n \) and each shared factor corresponding to \( k \)-th group of individuals we can define \( \delta_{nk} \), which equals 1 or 0 depending on relation type of the individual \( n \), which shares or does not, the factor \( k \). For a pedigree having \( N \) measured individuals and described by \( K \) factors with variances \( \sigma_{C,k}^2 \), shared by appropriate \( K \) overlapping groups, we have

\[ LH = \sum_{j=-J}^{J} \left( \prod_{k} \delta_k \right) \left\{ \prod_{n} \left\{ \exp \left\{ -\frac{(x_n - F(y_n) - \sum \delta_{nk} j\beta_k)^2}{2\sigma_{nk}^2} \right\} \left( \sigma_C \sqrt{\prod} \right) \right\} \right\} \]  

(3)

where indexes \( j \) change from \( -J \) to \( J \), \( p_j \) is the probability of definite shift \( j\beta_k \), corresponding to factor number \( k \); \( \delta_{nk} \) equals 1 or 0 depending on belonging to group \( k \); \( x_n \) is individual trait value; \( f(y) \) is predicted by covariates individual value; because all factors are supposed to be orthogonal \( \sigma_{nk} = \sigma_{nk} - \sigma_{cov} = \Sigma_k \delta_k \sigma_{k} \).

If the sample consists of MZ and DZ twin pairs, we may estimate additive genetic \( \sigma_{Add}^2 \) and common sibs (twins) environment \( \sigma_{Sib}^2 \) components, using LH (2), where for MZ families \( \sigma_{CMZ} = \sigma_{Add}^2 + \sigma_{Sib}^2 \), \( \sigma_{RMZ} = \sigma_{Tot}^2 - \sigma_{Covar} - \sigma_{Add}^2 - \sigma_{Sib}^2 \), for DZ families \( \sigma_{CDZ} = 0.5 \sigma_{Add}^2 + \sigma_{Sib}^2 \), \( \sigma_{RDZ} = \sigma_{Tot}^2 - \sigma_{Covar} - 0.5 \sigma_{Add}^2 \cdot \sigma_{Sib}^2 \). If the trait X is a liability of affection status B with affection threshold \( \tau \), the probability density LH (2) or (3) can be integrated independently on each \( X_i \) above or below the liability threshold, depending on affection status B. The one-dimensional integral is a tabulated function. The MLE parameters have in this model version two constraints. First is the liability variance in the sample \( \sigma_{Tot}^2 = 1 \); second the sample mean of liability equals to zero. So, in this case the individual specific variance component \( V_{RS} \) and the constant, included in the covariate adjustment \( F(Y_{ij}) \), are not independent parameters. The threshold parameter \( \tau \) is an estimated parameter.

In conventional formulation of VCA LH, parameters describing VC parameters \( \sigma_{C}^2 \cdot (\sigma_{Add}^2, \sigma_{Sib}^2) \) are included directly in the covariance matrix of normal distribution and automatically correspond to variance. In our LH expression, the part of the parameters that describe the VCs are included in the normal distribution as shifts of mean value (\( \beta \)). Suppose that we approximate the normally distributed shifts \( a \), corresponding to existing (simulated) shared variance \( \sigma_C^2 \), using stepwise distribution with only two categories: negative and positive shift \( \beta \), \( \beta \) with probability 0.5 each. In the MLE parameter \( \beta \), describing shift of mean, turns to achieve the mean value of the appropriate distribution category. After maximization we get the value of \( \beta \) as mean value of the positive part of the normal distribution of \( a; \beta^2 = 2\sigma_{Add}^2 \cdot 2\sigma_{C}^2 / \pi \). So the variance described with the parameter \( \beta \) in our model is less than \( \sigma_{C}^2 \cdot (2\sigma_{C}^2 / \pi) < \sigma_{C}^2 \). With such a distribution model we get a poor estimate of assumed VC. Additionally, the unexplained part of the true shared variance \( \sigma_C^2 \), which we tried to describe, will disturb the estimates of other parameters. If we suppose normality in the investigated trait, the range of summation indexes (number of discreet categories) and probabilities \( p_j \) should be selected to approximate the supposed normal distribution with sufficient precision. It is clear that for computation speed the minimal number of discreet categories is preferable. So the first step was to find the appropriate set of shifts and to test that the selected model (with synchronized shifts of mean value instead of correlation matrix) works for normally distributed traits. Thus, parameter estimations in the MLE procedure should be close to those of conventional VCA. Then we extended the model to binary traits (affection status).
Because the set of estimated parameters in our model (for a quantitative trait) is the same as in VCA, but they are differently included in the proposed LH, we denote our procedure as Quasi VCA.

For bivariate analysis of two quantitative traits, we used additional assumptions. For each type of shared factor included in the univariate analysis, we supposed the existence of decomposition to three independent factors, two trait-specific and one common for two traits; each is normally distributed. For each trait, the univariate factor is a linear combination of its specific part and the common factor. Hence for each type of factor included in the model, we have three summation indexes \( J \); two for independent, specific for each trait, sources of variability and one for the common source of variability for both traits. For individual specific variability, we also suppose the common source of variability for two traits, but here only this common factor is indexed independently for each individual. In bivariate analysis we use estimates of VC for both traits, which we obtained separately for each trait in univariate analyses. The parameters, which are to be found in MLE, are proportions of common variance for each trait \( \nu_{1c}, \nu_{2c} \) in each VC. The correlation coefficient in each type of VC is \( R_{C}^{2} = \nu_{1C} \times \nu_{2C} \). The sign of the correlation coefficient is included in \( R \) and can be different for different types of VC (for example, genetic additive and sibling environment components). The probability density LH for two quantitative traits is expressed (similar to LH (3)) as a sum through all shift indexes of the products of \( 2N \) one-variable normal distributions (each trait for each measured individual). One or both traits can be a liability \( X \) with threshold \( \tau \) of dichotomous affection status \( B \). In this case, the appropriate integration of probability density LH on individual liability traits \( X_{n} \), corresponding to affection status \( B_{n} \), should be included and \( \tau \) (or two different \( \tau_{1}, \tau_{2} \)) is an additional estimated parameter.

*Received*: 17 February 2014

*Accepted*: 30 April 2014